

Genome-Wide Population Structure Analyses of Three Minor Millets: Kodo Millet, Little Millet, and Proso Millet

Matthew Johnson*, Santosh Deshpande, Mani Vetriventhan, Hari D. Upadhyaya, and Jason G. Wallace

M. Johnson, J.G. Wallace, Johnson Institute of Plant Breeding, Genetics, and Genomics, Univ. of Georgia, 111 Riverbend Rd. Athens, GA; S. Deshpande, M. Vetriventhan, H.D. Upadhyaya, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, Telangana, India. MJ and JGW were supported by The Univ. of Georgia; SD, MV, and HDU were supported by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT).

ABSTRACT Millets are a diverse group of small-seeded grains that are rich in nutrients but have received relatively little advanced plant breeding research. Millets are important to smallholder farmers in Africa and Asia because of their short growing season, good stress tolerance, and high nutritional content. To advance the study and use of these species, we present genome-wide marker datasets and population structure analyses for three minor millets: kodo millet (*Paspalum scrobiculatum* L.), little millet (*Panicum sumatrense* Roth), and proso millet (*Panicum miliaceum* L.). We generated genome-wide marker data sets for 190 accessions of each species with genotyping-by-sequencing (GBS). After filtering, we retained between 161 and 165 accessions of each species, with 3461, 2245, and 1882 single-nucleotide polymorphisms (SNPs) for kodo, little, and proso millet, respectively. Population genetic analysis revealed seven putative subpopulations of kodo millet and eight each of proso millet and little millet. To confirm the accuracy of this genetic data, we used public phenotype data on a subset of these accessions to estimate the heritability of various agronomically relevant phenotypes. Heritability values largely agree with the prior expectation for each phenotype, indicating that these SNPs provide an accurate genome-wide sample of genetic variation. These data represent one of first genome-wide population genetics analyses—and the most extensive—in these species and the first genomic analyses of any sort for little millet and kodo millet. These data will be a valuable resource for researchers and breeders trying to improve these crops for smallholder farmers.

Abbreviations: GBS, genotyping-by-sequencing; ICRISAT, International Crops Research Institute for the Semi-Arid Tropics; SNP, single-nucleotide polymorphism; SSR, simple-sequence repeat.

CORE IDEAS

- Developed genome-wide SNP marker data for kodo, proso, and little millet
- Marker data used to analyze genetic diversity
- Heritability results of various traits used to validate genetic data

IT IS ESTIMATED that there are up to 7000 cultivated crop species in the world (Khoshbakht and Hammer 2008), yet major breeding and research efforts have focused on just a small number of these (Hammer et al., 2001). Many crops that have been ignored by modern research are still essential to the local communities that have relied on them for thousands of years, and they have potential for diversifying cropping systems around the world (Naylor et al., 2004). These crops are also genetic resources to increase global food security as climate changes and resources (land, water, fertilizers) become more limited. The ability to generate inexpensive, genome-wide data can quickly bring some of these crops into the modern genomics era (Varshney et al., 2012).

Millets are a diverse group of small-seeded grains that have been largely overlooked by modern genetics research. Although pearl millet [*Pennisetum glaucum* (L.) R. Br. syn. *Cenchrus americanus* (L.) Morrone],

Citation: Johnson, M., S. Deshpande, M. Vetriventhan, H.D. Upadhyaya, and J.G. Wallace. 2019. Genome-wide population structure analyses of three minor millets: Kodo millet, little millet, and proso millet. *Plant Genome* 12:190021. doi: 10.3835/plantgenome2019.03.0021

Received 14 Dec. 2018. Accepted 23 June 2019.

*Corresponding author (msj80897@uga.edu).

© 2019 The Author(s). This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Table 1. Genetic resources: Total ICRISAT accessions (<http://genebank.icrisat.org/Default>, accessed November 2018); proso millet core collection (Upadhyaya et al., 2011), kodo and little millet (Upadhyaya et al., 2014); ploidy information (Upadhyaya et al., 2011,2014); and estimated genome size for proso millet (Kubesova et al., 2010).

| Species | Total ICRISAT accessions | Accessions in ICRISAT core collection | Expressions sequence tags | Complete coding sequences | Ploidy level | Estimated genome size |
|---------------|--------------------------|---------------------------------------|---------------------------|---------------------------|----------------|-----------------------|
| Proso millet | 849 | 106 | 195 | 7 | $2n = 4x = 36$ | Mbp 1020.5 |
| Kodo millet | 665 | 75 | 29 | 0 | $2n = 4x = 40$ | unknown |
| Little millet | 473 | 55 | 12 | 0 | $2n = 4x = 36$ | unknown |

foxtail millet [*Setaria italica* (L.) P. Beauv.], and finger millet [*Eleusine coracana* (L.) Gaertn.] all have complete genome sequences and well-established germplasm resources (Varshney et al., 2017; Zhang et al., 2012; Prasad 2017; Sehgal et al., 2015; Hittalmani et al., 2017; Jia et al., 2013; Bennetzen et al., 2012; Hatakeyama et al., 2017), most other millets have few if any resources available (Goron and Raizada, 2015). These crops are often important for smallholder farmers, especially in south-eastern Asia and sub-Saharan Africa (FAO; <http://www.fao.org/docrep/W1808E/w1808e0e.htm>).

Proso millet, little millet, and kodo millet are three minor millets with limited modern genetic resources (Table 1), although proso millet has recently seen a significant increase in genomic studies (e.g., Rajput and Santra [2016], Habiyaemye et al. [2017], and Yue et al., [2016]), including a reference genome (Zou et al., 2019). Various germplasm repositories maintain collections of these species (Goron and Raizada, 2015). The current study focuses on those held by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), which maintains collections with 849, 473, and 665 accessions of proso, little, and kodo millet, respectively. These collections have been assessed for morphological and agronomic traits, and representative core collections have been created for each of them (Upadhyaya et al., 2011,2014). These millets are hardy C_4 grasses (Upadhyaya et al. 2014; Brown, 1999) with nutritional content on par with or superior to the major grains (Vetriventhan and Upadhyaya, 2018; Mengesha, 1966; Saleh et al., 2013; Kalinova and Moudry 2006).

Proso Millet

Proso millet is believed to have been domesticated ~10,000 yr ago. There have been multiple centers of origins proposed: northwestern China (Bettinger et al., 2007, 2010a,b), central China (Lu et al., 2009), and inner Mongolia (Zhao, 2005), with the most recent evidence pointing toward either one domestication event in China or one in China and one in Europe (Hunt et al., 2011). It is the third-oldest cultivated cereal after wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) (Habiyaemye et al., 2017; Upadhyaya et al., 2011). Proso millet is valued for its low water requirements (330 mm) and short growing season (60 d) (Habiyaemye et al., 2017; Shananhan et al., 1988). Proso millet varieties are classified into five races based on inflorescence morphology: miliaceum, patentissimum, contractum, compactum, and ovatum (de Wet et al., 1983). Proso millet is tetraploid ($2n = 4x = 36$)

(Saha et al., 2016) with an allotetraploid origin (Hunt et al., 2014). Proso millet has historically be the most widely grown of the three millets studied here, with cultivation concentrated in the former Soviet Union and India (Roshevits, 1980) and with significant production in the United States (Baltensperger, 2002).

Proso millet genetic diversity has been investigated with a variety of genetic markers; however, most of them at a very small scale (<100 markers; reviewed in Habiyaemye et al. [2017]). Recently, however, the most extensive genetic analysis in proso millet identified over 400,000 SNP markers and 35,000 simple-sequence repeats (SSRs) from the transcriptomes of two proso millet accessions (Yue et al., 2016). Population-level analyses were made by Rajput and Santra (2016) using 100 SSRs and 90 proso millet accessions; it was found that there were some connections between genetic clustering and geographic origin. There also has been quantitative trait loci mapping using 833 SNPs from GBS data (Rajput and Santra, 2016) and a just-published reference genome (Zou et al., 2019).

Kodo Millet

Kodo millet was domesticated in India ~3000 yr ago, and India has historically been the major center of cultivation (de Wet et al., 1983). Kodo millet accessions have been classified into three races based on panicle morphology: regularis, irregularis, and variabilis (de Wet et al., 1983; Prasada Rao et al., 1993). Kodo millet is tetraploid ($2n = 4x = 40$) (Saha et al., 2016) and it is valued for its ability to produce consistently in hot, drought-prone arid and semiarid land (Dwivedi et al., 2012). A few sets of molecular markers have been developed for kodo millet based on random-amplification-of-polymorphic-DNA (RAPD) markers (M'Ribu and Hilu 1996), gene-specific primer sets (Kushwaha et al., 2014), and semitargeted polymerase chain reaction amplification (Yadav et al., 2016); the latter study concluded that the 96 accessions they used could be divided into four groups that showed little connection between geographic region and genetic relationship of the accessions. There have been no truly genome-wide datasets on this species before now.

Little Millet

Little millet was domesticated ~5000 yr ago in India (de Wet et al., 1983). It has historically been grown mainly in India, Myanmar, Nepal, and Sri Lanka (Prasada Rao et al., 1993). Little millet accessions have been classified into two races based on panicle morphology, nana and robusta, with two subraces per race (laxa and erecta for nana and laxa

and compacta for robusta) (de Wet et al., 1983; Prasada Rao et al., 1993). Little millet is tetraploid ($2n = 4x = 36$) (Saha et al., 2016). Like kodo millet, little millet can give consistent yields on marginal lands in drought-prone arid and semi-arid regions, and it is an important crop for regional food stability (Dwivedi et al., 2012). Little millet is arguably the least studied of these three millets, and we are unaware of any molecular markers developed for it outside of specific single genes (Goron and Raizada 2015) and a small set of RAPD markers whose details were not described (M.S. Swaminathan Research Foundation, 2000).

Expanding Genetic Resources

As mentioned above, the genetic and genomic resources of proso millet, kodo millet, and little millet are very limited (Saha et al., 2016; Goron and Raizada 2015), although the situation for proso millet, at least, is improving. The main focus of the resources have been centered on the accessions stored in gene banks, expression sequence tags, and complete coding sequences.

Core collections for each of these species were created several years ago consisting of ~10% of ICRISAT's collection for each species (Upadhyaya et al., 2011, 2014). These collections have been assessed for morphoagronomic traits but no genetic data has been available for them. Our goal in this project was to generate genome-wide marker data on each of these species to enable population genetic analysis and empower breeders and researchers to make more informed decisions about germplasm selection and representation. We genotyped 190 accessions of each species, including the entire core collections of each using GBS (Elshire et al., 2011) to generate the most comprehensive population genetics resource for each of these species to date. These data dramatically expand the genomic resources available to each of these crops and will help make better use of them moving forward.

MATERIALS AND METHODS

Plant Materials

One-hundred ninety accessions (Supplemental Data) were taken from the ICRISAT genebank for each millet species. These included the accessions from the core collection for proso millet, kodo millet, and little millet (Upadhyaya et al., 2011, 2014). The remaining samples were chosen to broadly sample the available diversity based on the cluster information used to create the core collections.

DNA Extraction and Sequencing

Seedlings of each accession were grown at the ICRISAT research station in Patancheru, India, in 2015. DNA was extracted from 55-d-old plants, with one plant from each accession used for extraction, by the modified CTAB method (Mace et al., 2003), lyophilized, and shipped to the Genomic Diversity Facility at Cornell University for GBS (Elshire et al., 2011). The GBS library preparation followed standard methods (Wallace and Mitchell 2017) using the PstI restriction enzyme. For kodo millet, 190

samples plus two blanks were multiplexed into a single lane for sequencing on an Illumina HiSeq 2500 with single-end 100-bp sequencing. For little and proso millet, the procedure was similar except that samples were multiplexed into two lanes of 95 samples plus one blank. The raw sequence data for these samples is available at the Sequence Read Archive, accessions PRJNA494158.

Single-Nucleotide Polymorphism Calling

Single-nucleotide polymorphisms were called from raw sequencing data using the TASSEL (Bradbury et al., 2007) GBS v2 SNP pipeline. Since this pipeline uses alignment to a reference genome and none of these species had such a reference when we performed the analyses, we included slight alterations to make use of the UNEAK filter (Lu et al., 2013) for reference-free alignment of tags to each other. All code for this analysis is available in the Supplemental data files and on GitHub (https://github.com/wallacelab/2018_minor_millet).

Filtering

Low-quality SNPs were removed by filtering the raw genotype data to remove sites with >50% missing data, minor allele frequency <5%, and >20 to 25% heterozygosity. (The heterozygosity filter removes false SNPs as a result of paralogs misaligning in polyploid species (Wallace and Mitchell, 2017). Samples with >50% missing data across all remaining sites were also filtered out.

The above filters reduced total missing data to <20% in each species (9% in kodo, 16.8% in little, 13% in proso). We tested imputation with LinkImpute (Money et al., 2015) to reduce this further; however, the imputed data did not result in any appreciable differences (Supplemental Fig. S1–S3) in downstream analyses, so all analyses were conducted with the unimputed data.

Phylogeny

Phylogenetic networks were constructed using the NeighborNet method (Bryant and Moulton, 2003) in SplitsTree4 V4.14.4 (Huson and Bryant, 2005). Genotype data was converted to SplitsTree-compatible NEXUS files by first using TASSEL version 5.2.29 (Bradbury et al., 2007) to export as a PHYLIP (interleaved) format, which was then converted to NEXUS format using Alter (Glez-Peña et al., 2010) (<http://www.sing-group.org/ALTER/>). The resulting files were manually edited to change the data type to dna and replace all colons (:) with underscores (_), at which point the files were loaded into SplitsTree for network creation.

Population Structure Determination

Population structure analysis was performed with FastSTRUCTURE v1.0 (Raj et al., 2014), with the number of potential populations (k) varying from 1 to 15. The optimum population size was determined by chooseK and results were visualized with Distruct, both parts of the FastSTRUCTURE software package. Default parameters were used for all programs. Samples were assigned to a

population if they had at least 60% membership in that population (this cutoff is arbitrary, but using a higher cutoff did not appreciably change population membership [data not shown]). Fixation index values were calculated for comparisons between all populations in each species as additional validation (Supplemental Data)

Plotting Accessions by Geographic Data

Geographic data on kodo millet and little millet was plotted in Python using basemap (Hunter 2007) and the known GPS coordinates, then colored by subpopulation. Since proso millet had only the country of origin for some accessions, its map was created manually in Inkscape 0.91 (<https://inkscape.org>) by placing dots (colored by subpopulation) in the country of origin for each accession.

Principal Coordinate Analysis

Genetic principal coordinates were calculated using multidimensional scaling analysis in TASSEL V5.2.29 (Bradbury et al., 2007).

Heritability Analysis

Phenotype data for heritability analysis was taken from public data on the kodo and little millet core collections (Upadhyaya et al., 2014), and proso millet phenotype data was provided by collaborators at IRCISAT (Supplemental Data). Phenotypes were fit as part of a mixed linear model in TASSEL (Bradbury et al., 2007) with a kinship matrix as the only covariate using default parameters. Narrow-sense heritability (h^2) was estimated as the ratio of genetic variance to total variance in the model.

Software

The following software and packages were used as part of this analysis:

- SplitsTree4: V4.14.4 (Huson and Bryant, 2005)
- Faststructure: 1.0: <https://rajanil.github.io/fastStructure/> (Raj et al., 2014)
- TASSEL 5.2.29: <http://www.maizegenetics.net/tassel> (Bradbury et al., 2007)
- PLINK 2.0: <http://zzz.bwh.harvard.edu/plink/> (Purcell et al., 2007)
- Inkscape 0.91: <https://inkscape.org/release/inkscape-0.91/?latest=1>
- GNU Parallel 20141022 (Tange, 2011)
- Matplotlib 2.2.2 (Hunter, 2007)
- Pandas 0.17 (McKinney 2010)
- NumPy 1.11.0 (Oliphant, 2006)
- Python 3.5.2
- VCF Tools 0.1.13 (Danecek et al., 2011)

Table 2. Number of single nucleotide polymorphisms (SNPs) and accessions before and after filtering.

| Species | Raw SNPs | Filtered SNPs | Raw accessions | Filtered accessions |
|---------------|----------|---------------|----------------|---------------------|
| Kodo millet | 12995 | 3461 | 190 | 165 |
| Little millet | 14473 | 2245 | 190 | 165 |
| Proso Millet | 12839 | 1882 | 190 | 161 |

RESULTS AND DISCUSSION

Genome-Wide Marker Sets

Genotype data was generated using GBS (Elshire et al., 2011) with PstI restriction digestion. This resulted in ~12,000 to 14,000 raw SNPs per species (Table 2), although many of these are probably artifacts resulting from misalignments. Low-quality SNPs were removed by filtering the raw genotype data to remove sites with >50% missing data, minor allele frequency <5%, and >20 to 25% heterozygosity. (The heterozygosity filter removes false SNPs resulting from paralogs misaligning in polyploid species (Wallace and Mitchell, 2017)). Samples with >50% missing data across all remaining sites were also filtered out. The final, filtered genotype data (Supplemental Data) consists of 3461 SNPs across 165 accessions of kodo millet, 2245 SNPs across 165 accessions of little millet, and 1882 SNPs across 161 accessions of proso millet. (Table 2)

Proso Millet

Population structure analysis of proso millet with fastSTRUCTURE (Raj et al., 2014) grouped the 161 postfiltering samples into eight putative subpopulations (Fig. 1). Some of the subpopulations consist almost entirely of pure individuals of that subpopulation, implying strong separation from the other subpopulations (e.g., Groups 7 and 8), while others (e.g., Groups 2–5) show significant admixture among the subpopulations. We observed that small changes in fastSTRUCTURE parameters would significantly alter the division and number of subpopulations in this admixed group (data not shown), further indicating that the divisions among the highly admixed subpopulations are weak and should be interpreted with caution.

A phylogenetic network of these samples (Fig. 1) mirrors this, where most of the subpopulations group together, although Subpopulations 6 and 8 are strongly separated from the rest. Divisions among the main group of populations appear weak, as evidenced by the large number of alternative splits (webbing) in the phylogenetic web and the fact that two populations (2 and 4) are split into two to three groups across the phylogeny. The genetic principal coordinates of these samples (Fig. 2) indicate a similar pattern, where most samples cluster together but with Subpopulations 6 and 8 distinct.

While the race data was not available for proso millet in this study, there is strong evidence that race is not a good indicator of genetic relatedness among accessions (Vetriventhan and Upadhyaya 2018). Existing core collections were designed with the races of the species being a central focus. Core collections likely could be improved

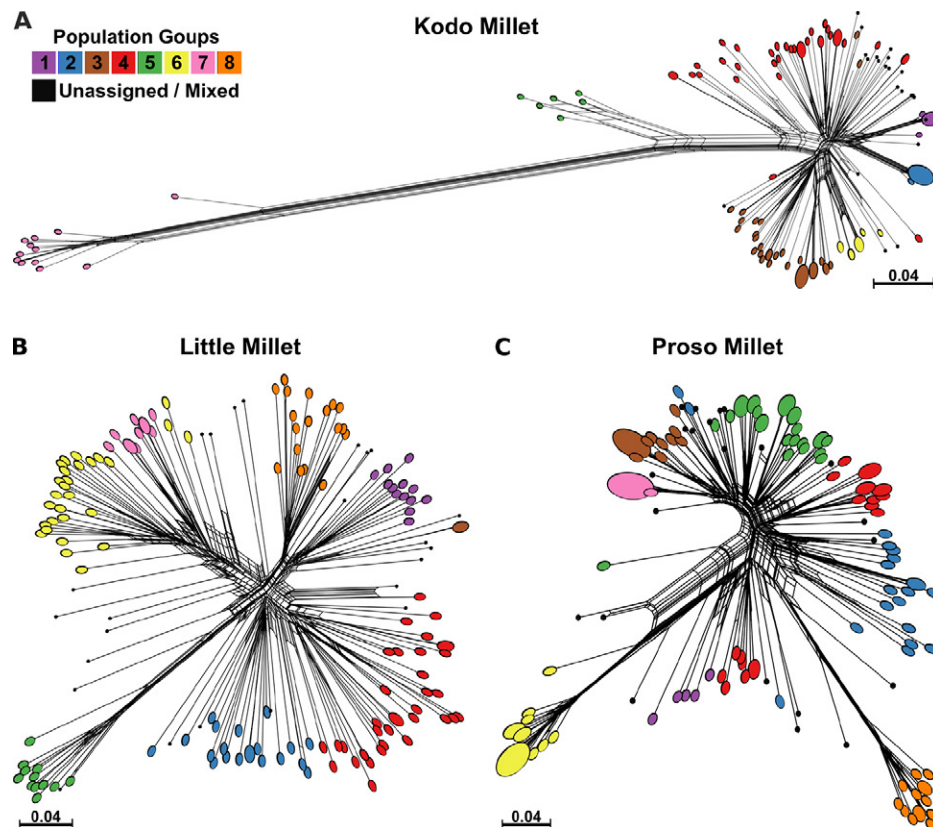


Fig. 1. Phylogenetic trees for (A) kodo millet, (B) little millet, and (C) proso millet colored by population calls from fastSTRUCTURE for accessions that fit into a population with at least 60% fit. Each accession has one line that ends in a point and has branching to show alternative branching options. Where multiple points of the same population grouped closely together, a single oval with a size proportional to the number of accession it encompasses was used.

by using population structure and genomic data to improve the genetic diversity of the collections.

The plotting of the 109 accessions with known geographic location supports the population analysis (Supplemental Fig. S4).

The reference genome was published after the bulk of the work for this paper had been done, and when analyses were reran using the reference genome, there were no significant difference found from the original analysis (Supplemental Fig. S5; the SNPs called from the new reference genome were included as well in Supplemental Data)

Kodo Millet

Population structure analysis places the 165 kodo millet samples into seven putative subpopulations (Fig. 2). Most accessions cluster together, but two subpopulations (5 and 7) are strongly separated from the rest (Fig. 1,2). The separation of Subpopulation 7 from the other samples drives the largest genetic differences in this species, with 67.5% of the genetic variation along this axis (PC1 in Fig. 2).

These patterns of population structure do not correlate with existing race designations (Supplemental Fig. S6), implying that existing race designations do not strongly correlate to genetic groupings. Similar to proso millet, this implies that core collections of kodo millet could be improved by using population and genomic data instead of race designation.

Only 17 of the kodo millet accessions had known geographic origins, all of them in India. Plotting these origins geographically supports the population analysis by showing clustering of Population 4 (Supplemental Fig. S7), although more samples would be needed to completely confirm this interpretation.

Little Millet

The 165 filtered accessions of little millet were grouped into eight putative subpopulations by fastSTRUCTURE (Fig. 2). Little millet appears to have weaker population structure than the other two species (compare the variance explained by each principal coordinate, Fig. 2). Subpopulation 5 is relatively well separated from the others, followed by Subpopulations 6 and 7, and then the other subpopulations being relatively close together. (These subpopulation designations would shift around with minor changes in fastSTRUCTURE parameters, implying they are very weakly separated [data not shown].) Similar to the other two species, little millet population structure did not correlate with existing race designations (Supplemental Fig. S8). Plotting the 29 accessions with known geographic location supports the population analysis by showing distinct clustering of Populations 4 and 5 (Supplemental Fig. S9).

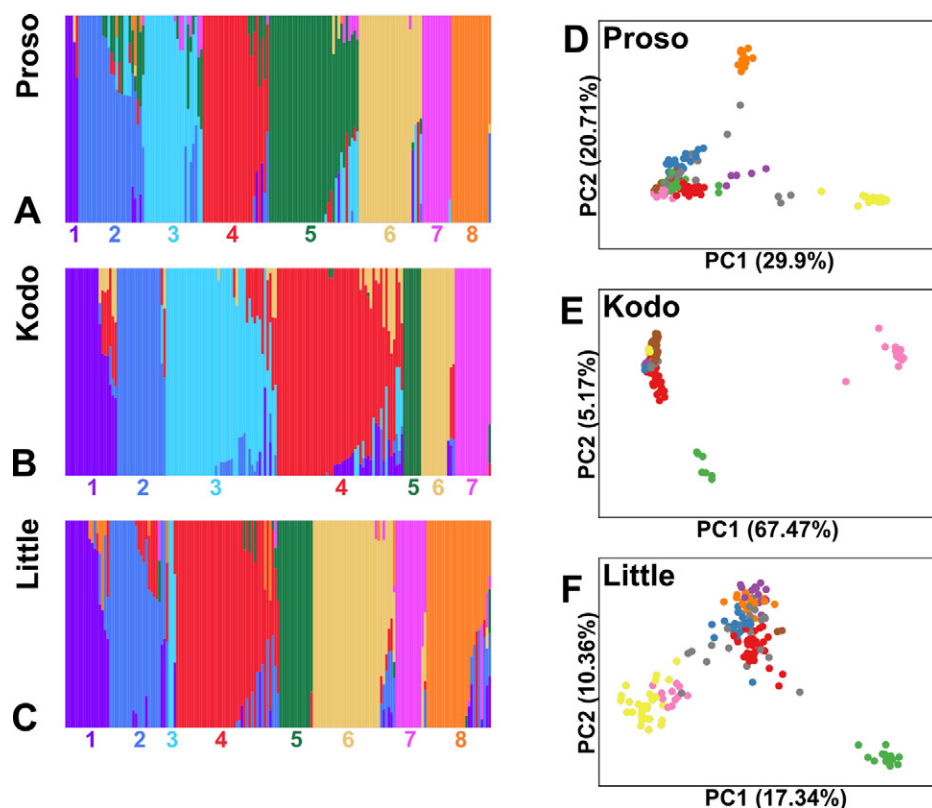


Fig. 2. (A–C) Visual representation of how well individual accessions fit into putative populations from Distruct, a part of the fastSTRUCTURE program. The vertical bars each represent a single accession, and the colors correspond to populations as listed on the bottom of each subfigure: (A) proso millet, (B) kodo millet, and (C) little millet. (D–F) Principle coordinate analysis based off of genetic distances and colored by putative populations for (A) proso millet, (B) kodo millet, and (C) little millet.

Validation of Genotype Data via Estimated Heritability

To validate these genetic data, we used public phenotype data on the kodo millet and little millet core collections (Upadhyaya et al., 2014) and proso millet phenotypic data that was provided by ICRISAT (Supplemental Data). These data were used to estimate narrow-sense heritability (h^2) for flowering time and plant height (Table 3). Our expectation was that true genotype data should result in moderate to high heritability values for some phenotypes, especially ones (such as flowering time and plant height) that are known to have a strong genetic component.

Heritability was estimated using a mixed linear model in TASSEL (see Materials and Methods section) to obtain estimated variance components. Little millet traits exhibited heritability from 0.209 to 0.807, while kodo millet traits were slightly lower, ranging from 0.0 to 0.505 (Supplemental Data). There were no other phenotypes provided for proso millet other than what was included in Table 3. In all three species, flowering time and plant height were some of the most heritable traits (Table 3). These traits are all known to be under strong genetic control in other grass crops (e.g., Buckler et al., 2009; Peiffer et al., 2014; Mace et al., 2013; Morris et al., 2012; Ma et al., 2016; Alqudah and Schnurbusch, 2014). Since poor-quality SNPs should show little to no relationship with phenotype, these results imply that the SNP datasets we have generated accurately represent the

genetic variation within these populations and can be used for real-world breeding applications.

CONCLUSIONS

The results from this study represent one of the first genome-wide analyses for proso millet and the first for kodo millet and little millet, three orphan crops that are important for food security in developing nations. We have identified thousands of SNPs in each of these species that accurately capture the population structure of each, as indicated by the geographic correlations and estimates of narrow-sense heritability. Our analyses can be used as a foundation for further exploration into the genetics of these species including selecting appropriate breeding materials and identifying priority populations for further collection and curation.

Existing core collections were designed with the races of the species being a central focus. The evidence strongly implies that race is not a good determiner of genetic relatedness, and, as such, the core collections likely could be improved by using modern genomic data to improve the genetic diversity of the collections.

For both major and minor crops, obtaining genetic data is now (almost) trivial, even in species with highly complex genomes and no prior history of genetic analysis. As the price of DNA sequencing continues to drop, more and more orphan species will have genotype data

Table 3. Estimated heritabilities for minor millets.

| Millet species | Trait | Heritability |
|----------------|-----------------------|--------------|
| Little | Days to 50% flowering | 0.807 |
| Little | Plant height (cm) | 0.664 |
| Proso | Days to 50% flowering | 0.69 |
| Proso | Plant height (cm) | 0.615 |
| Kodo | Days to 50% flowering | 0.505 |
| Kodo | Plant height (cm) | 0.293 |

available. The major question going forward will be how to best deploy these data to benefit breeders and growers. Given that the price of phenotyping is often the limiting factor in many studies (Cobb et al., 2013), finding ways to deploy genomic prediction and high-throughput phenotyping for orphan crops likely will be the next major step to democratize modern genomics for the developing world.

Supplemental Information

Supplemental Fig. S1. Principal coordinate analysis for kodo millet using imputed genotype data. Population structure and groupings are almost identical to those made with unimputed data (main text Fig. 2E).

Supplemental Fig. S2. Principal coordinate analysis for little millet using imputed genotype data. Population structure and groupings are almost identical to those made with unimputed data (main text Fig. 2D).

Supplemental Fig. S3. Principal coordinate analysis for kodo millet using imputed genotype data. Population structure and groupings are almost identical to those made with unimputed data (main text Fig. 2F).

Supplemental Fig. S4. Plotting of Proso millet accessions by country of origin. Within-country coordinates were not available, so points within each country are displayed as a grid. Points are colored by subpopulation (see main text Fig. 2); black indicates the accession didn't fit into any given population.

Supplemental Fig. S5. Principal Coordinate analysis of Proso millet using the recently published reference genome (Zou et al., 2019). These results are consistent with those done without a reference genome (main text Fig. 2D).

Supplemental Fig. S6. PC analysis of kodo millet using race to color the accessions. Races do not correspond with genetic clusters, indicating that they do not reflect genetic relationships

Supplemental Fig. S7. Kodo millet accessions plotted by GPS coordinates. Only a few kodo millet accessions had GPS data available, though those that do are relatively consistent with subpopulations clustering by region of origin. Points are colored by subpopulation (see main text Fig. 2); black indicates the accession didn't fit into any given population.

Supplemental Fig. S8. PC analysis of Little millet using race to color the accessions. Race does not show the diversity of genetic clusters, and as such it is not a good reflection of genetic relationships.

Supplemental Fig. S9. Little millet accessions plotted by GPS coordinates. Only a few little millet accessions had GPS data available, though those that do are relatively consistent with subpopulations clustering by region of origin. Points are colored by subpopulation (see main text Fig. 2); black indicates the accession didn't fit into any given population.

Author Contributions

Santosh Deshpande, Mani Vetriventhan, and Hari Upadhyaya grew the plants and extracted the DNA. Jason Wallace called the SNPs. Matthew Johnson cleaned and analyzed the data. Matthew Johnson wrote the first draft of the paper and all authors contributed to edits.

Conflicts of Interest

The authors declare that they have no known conflicts of interests.

REFERENCES

- Alqudah, A.M., and T. Schnurbusch. 2014. Awn primordium to tipping is the most decisive developmental phase for spikelet survival in barley. *Funct. Plant Biol.* 41:424. doi:10.1071/FP13248
- Baltensperger, D.D. 2002. Progress with proso, pearl and other millets. In: J. Janick and A. Whipkey, editors, *Trends in new crops and new uses*. ASHS Press, Alexandria, VA. p. 100–103.
- Bennetzen, J.L., J. Schmutz, H. Wang, R. Percifield, J. Hawkins, A.C. Pontaroli, et al. 2012. Reference genome sequence of the model plant *Setaria*. *Nat. Biotechnol.* 30:555–561. doi:10.1038/nbt.2196
- Bettinger, R.L., L. Barton, and C. Morgan. 2010a. The origins of food production in north China: A different kind of agricultural revolution. *Evolutionary Anthropology: Issues, News Rev. (Melb.)* 19:9–21.
- Bettinger, R.L., L. Barton, C. Morgan, F. Chen, H. Wang, T.P. Guilderson, et al. 2010b. The transition to agriculture at Dadiwan, People's Republic of China. *Curr. Anthropol.* 51:703–714. doi:10.1086/655982
- Bettinger, R.L., L. Barton, P.J. Richerson, R. Boyd, H. Wang, and W. Choi. 2007. The transition to agriculture in northwestern China. *Dev. Quat. Sci.* 9:83–101. doi:10.1016/S1571-0866(07)09008-2
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. doi:10.1093/bioinformatics/btm308
- Brown, R.H. 1999. Agronomic implications of C4 photosynthesis. *C4 plant biology*: 473–507.
- Bryant, D., and V. Moulton. 2003. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* 21:255–265. doi:10.1093/molbev/msh018
- Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, et al. 2009. The genetic architecture of maize flowering time. *Science* 325:714–718. doi:10.1126/science.1174276
- Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* 126:867–887. doi:10.1007/s00122-013-2066-0
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. doi:10.1093/bioinformatics/btr330
- de Wet, J.M.J., D.E. Brink, K.E.P. Rao, and M.H. Mengesha. 1983. Diversity in Kodo millet, *Paspalum scrobiculatum*. *Econ. Bot.* 37:159–163. doi:10.1007/BF02858779
- Dwivedi, S., H. Upadhyaya, S. Senthilvel, C. Hash, K. Fukunaga, X. Diao, et al. 2012. Millets: Genetic and genomic resources. *Plant Breed. Rev.* 35:247–375. doi:10.1002/9781118100509.ch5

- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi:10.1371/journal.pone.0019379
- Glez-Peña, D., D. Gómez-Blanco, M. Reboiro-Jato, F. Fdez-Riverola, and D. Posada. 2010. ALTER: Program-oriented conversion of DNA and protein alignments. *Nucleic Acids Res.* 38:W14–W18. doi:10.1093/nar/gkq321
- Goron, T.L., and M.N. Raizada. 2015. Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. *Front. Plant Sci.* 6. doi:10.3389/fpls.2015.00157
- Habiyaremye, C., J.B. Matanguihan, J. D'Alpoim Guedes, G.M. Ganjyal, M.R. Whiteman, K.K. Kidwell, et al. 2017. Proso millet (*Panicum miliaceum* L.) and its potential for cultivation in the Pacific Northwest, U.S.: A review. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01961
- Hatakeyama, M., S. Aluri, M.T. Balachadran, S.R. Sivarajan, A. Patrignani, S. Grüter, et al. 2017. Multiple hybrid *de novo* genome assembly of finger millet, an orphan allotetraploid crop. *DNA Res.* 25:39–47. doi:10.1093/dnares/dsx036
- Hammer, K., J. Heller, and J. Engels. 2001. Monographs on underutilized and neglected crops. *Genet. Resour. Crop Evol.* 48:3–5. doi:10.1023/A:1011253924058
- Hittalmani, S., H.B. Mahesh, M.D. Shirke, H. Biradar, G. Uday, Y.R. Aruna, et al. 2017. Genome and transcriptome sequence of finger millet (*Eleusine coracana* (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. *BMC Genomics* 18:465. doi:10.1186/s12864-017-3850-z
- Hunt, H.V., F. Badakshi, O. Romanova, C.J. Howe, M.K. Jones, and J.S. Heslop-Harrison. 2014. Reticulate evolution in *Panicum* (Poaceae): The origin of tetraploid broomcorn millet, *P. miliaceum*. *J. Exp. Bot.* 65:3165–3175. doi:10.1093/jxb/eru161
- Hunt, H.V., M.G. Campana, M.C. Lawes, Y.J. Park, M.A. Bower, C.J. Howe, et al. 2011. Genetic diversity and phylogeography of broomcorn millet (*Panicum miliaceum* L.) across Eurasia. *Mol. Ecol.* 20:4756–4771. doi:10.1111/j.1365-294X.2011.05318.x
- Hunter, J.D. 2007. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9:90–95. doi:10.1109/MCSE.2007.55
- Huson, D.H., and D. Bryant. 2005. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23:254–267. doi:10.1093/molbev/msj030
- Jia, G., X. Huang, H. Zhi, Y. Zhao, Q. Zhao, W. Li, et al. 2013. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* 45:957–961. doi:10.1038/ng.2673
- Kalinova, J., and J. Moudry. 2006. Content and quality of protein in proso millet (*Panicum miliaceum* L.) varieties. *Plant Foods Hum. Nutr.* 61:43. doi:10.1007/s11130-006-0013-9
- Khoshbakht, K., and K. Hammer. 2008. How many plant species are cultivated? *Genet. Resour. Crop Evol.* 55:925–928. doi:10.1007/s10722-008-9368-0
- Kubesova, M., L. Moravcova, J. Suda, V. Jarosik, and P. Pysek. 2010. Naturalized plants have smaller genomes than their non-invading relatives: A flow cytometric analysis of the Czech alien flora. *PRESLIA-PRAHA* 82:81–96.
- Kushwaha, H., K.W. Jillo, V.K. Singh, A. Kumar, and D. Yadav. 2014. Assessment of genetic diversity among cereals and millets based on PCR amplification using Dof (DNA binding with One Finger) transcription factor gene-specific primers. *Plant Syst. Evol.* 301:833–840. doi:10.1007/s00606-014-1095-8
- Lu, F., G.P. Copenhaver, A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, et al. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9:e1003215. doi:10.1371/journal.pgen.1003215
- Lu, H., J. Zhang, K.B. Liu, N. Wu, Y. Li, K. Zhou, et al. 2009. Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *Proc. Natl. Acad. Sci. USA* 106:7367–7372. doi:10.1073/pnas.0900158106
- Ma, X., F. Feng, H. Wei, H. Mei, K. Xu, S. Chen, et al. 2016. Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01801
- Mace, E.S., K.K. Buhariwalla, H.K. Buhariwalla, and J.H. Crouch. 2003. A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol. Biol. Report.* 21:459–460. doi:10.1007/BF02772596
- Mace, E.S., C.H. Hunt, and D.R. Jordan. 2013. Supermodels: Sorghum and maize provide mutual insight into the genetics of flowering time. *Theor. Appl. Genet.* 126:1377–1395. doi:10.1007/s00122-013-2059-z
- McKinney, W. 2010. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, Austin, TX. 28 June–3 July, 2010. p. 51–56.
- Mengesha, M.H. 1966. Chemical composition of teff (*Eragrostis tef*) compared with that of wheat, barley and grain sorghum. *Econ. Bot.* 20:268–273. doi:10.1007/BF02904277
- Money, D., K. Gardner, Z. Migicovsky, H. Schwaninger, G.Y. Zhong, and S. Myles. 2015. LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3: Genes, Genomes, Genet.* 5:2383–2390. doi:10.1534/g3.115.021667
- Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, et al. 2012. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. USA* 110:453–458. doi:10.1073/pnas.1215985110
- M'Ribu, H.K., and K.W. Hilu. 1996. Application of random amplified polymorphic DNA to study genetic diversity in *Paspalum scrobiculatum* L. (Kodo millet, Poaceae). *Genet. Resour. Crop Evol.* 43:203–210. doi:10.1007/BF00123272
- M.S. Swaminathan Research Foundation. 2000. Ninth annual report 1998–99. Centre for Research on Sustainable Agricultural and Rural Development. Chennai, India
- Naylor, R.L., W.P. Falcon, R.M. Goodman, M.M. Jahn, T. Sengooba, H. Tefera, et al. 2004. Biotechnology in the developing world: A case for increased investments in orphan crops. *Food Policy* 29:15–44. doi:10.1016/j.foodpol.2004.01.002
- Olipiant, T.E. 2006. A guide to NumPy, Trelgol Publishing, USA.
- Peiffer, J.A., M.C. Romay, M.A. Gore, S.A. Flint-Garcia, Z. Zhang, M.J. Milard, et al. 2014. The genetic architecture of maize height. *Genetics* 196:1337–1356. doi:10.1534/genetics.113.159152
- Prasad, M., editor. 2017. The foxtail millet genome. Springer, Cham, Switzerland. doi:10.1007/978-3-319-65617-5
- Prasada Rao, K., J. De Wet, V. Gopal Reddy, and M. Mengesha. 1993. Diversity in the small millets collection at ICRISAT. In: K.W. Riley, S.C. Gupta, A. Seetharam, and J.N. Mushonga, editors, *Advances in small millets*. Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi, India.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81:559–575.
- Raj, A., M. Stephens, and J.K. Pritchard. 2014. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197:573–589. doi:10.1534/genetics.114.164350
- Rajput, S.G., and D.K. Santra. 2016. Evaluation of genetic diversity of proso millet germplasm available in the United States using simple-sequence repeat markers. *Crop Sci.* 56:2401–2409. doi:10.2135/cropsci2015.10.0644
- Roshevits, R.Y. 1980. Grasses: An introduction to the study of fodder and cereal grasses. Springer-Verlag, Heidelberg. doi:10.2307/2806054
- Saha, D., M.V.C. Gowda, L. Arya, M. Verma, and K.C. Bansal. 2016. Genetic and genomic resources of small millets. *Crit. Rev. Plant Sci.* 35:56–79. doi:10.1080/07352689.2016.1147907
- Saleh, A.S., Q. Zhang, J. Chen, and Q. Shen. 2013. Millet grains: Nutritional quality, processing, and potential health benefits. *Compr. Rev. Food Sci. Food Saf.* 12:281–295. doi:10.1111/1541-4337.12012
- Sehgal, D., S.K. Parida, L. Skot, R. Singh, R.K. Srivastava, S.P. Das, et al. 2015. Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. *PLoS One* 10:e0122165. doi:10.1371/journal.pone.0122165
- Shanahan, J., R. Anderson, and B. Greb. 1988. Productivity and water use of proso millet grown under three crop rotations in the Central Great Plains. *Agron. J.* 80:487–492. doi:10.2134/agronj1988.00021962008000030019x

- Tange, O. 2011. GNU Parallel—The command-line power tool. The USENIX Magazine, February. p. 42–47.
- Upadhyaya, H., S. Dwivedi, S. Singh, S. Singh, M. Vetriventhan, and S. Sharma. 2014. Forming core collections in barnyard, kodo, and little millets using morphoagronomic descriptors. *Crop Sci.* 54:2673–2682. doi:10.2135/cropsci2014.03.0221
- Upadhyaya, H., S. Sharma, C. Gowda, V.G. Reddy, and S. Singh. 2011. Developing proso millet (*Panicum miliaceum* L.) core collection using geographic and morpho-agronomic data. *Crop Pasture Sci.* 62:383–389. doi:10.1071/CP10294
- Varshney, R.K., J.M. Ribaut, E.S. Buckler, R. Tuberosa, J.A. Rafalski, and P. Langridge. 2012. Can genomics boost productivity of orphan crops? *Nat. Biotechnol.* 30:1172–1176. doi:10.1038/nbt.2440
- Varshney, R.K., C. Shi, M. Thudi, C. Mariac, J. Wallace, P. Qi, et al. 2017. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.* 35:969–976. doi:10.1038/nbt.3943
- Vetriventhan, M., and H.D. Upadhyaya. 2018. Diversity and trait-specific sources for productivity and nutritional traits in the global proso millet (*Panicum miliaceum* L.) germplasm collection. *Crop J.* 6:451–463. doi:10.1016/j.cj.2018.04.002
- Wallace, J.G. and S.E. Mitchell. 2017. Genotyping-by-sequencing. *Curr. Protoc. Plant Biol.* 2:64–77. doi:10.1002/cppb.20042
- Yadav, Y., G.R. Lavanya, S. Pandey, M. Verma, C. Ram, and L. Arya. 2016. Neutral and functional marker based genetic diversity in kodo millet (*Paspalum scrobiculatum* L.). *Acta Physiol. Plant.* 38. doi:10.1007/s11738-016-2090-1
- Yue, H., L. Wang, H. Liu, W. Yue, X. Du, W. Song, et al. 2016. De novo assembly and characterization of the transcriptome of broomcorn millet (*Panicum miliaceum* L.) for gene discovery and marker development. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01083
- Zhang, G., X. Liu, Z. Quan, S. Cheng, X. Xu, S. Pan, et al. 2012. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* 30:549–554. doi:10.1038/nbt.2195
- Zhao, Z. 2005. Palaeoethnobotany and its new achievements in China. *Kaogu (Archaeology)* 7:e49.
- Zou, C., L. Li, D. Miki, D. Li, Q. Tang, L. Xiao, et al. 2019. The genome of broomcorn millet. *Nat. Commun.* 10:436. doi:10.1038/s41467-019-08409-5